



DATA
TERRA



CENTRES DE RÉFÉRENCE THÉMATIQUES

Systeme Terre-Environnement

Contribuer à la science ouverte

Par
Ghislaine Abbassi, responsable Communication Data Terra
et
Olivier Norvez, responsable Animation PNDB



Love Data Week

Nancy, Université de Lorraine, 16 mars 2023



O^S Ouvrir
la science !



SOMMAIRE



DATA
TERRA



Pôle National
de Données de Biodiversité

01

LES DÉFIS DE
DEMAIN

02

IRs Data Terra et PNDDB
QUI SOMMES-NOUS ?

03

CENTRES DE RÉFÉRENCE
THÉMATIQUES



01 LES DÉFIS DE DEMAIN



LA TERRE, UN SYSTÈME DYNAMIQUE COMPLEXE



DATA
TERRA

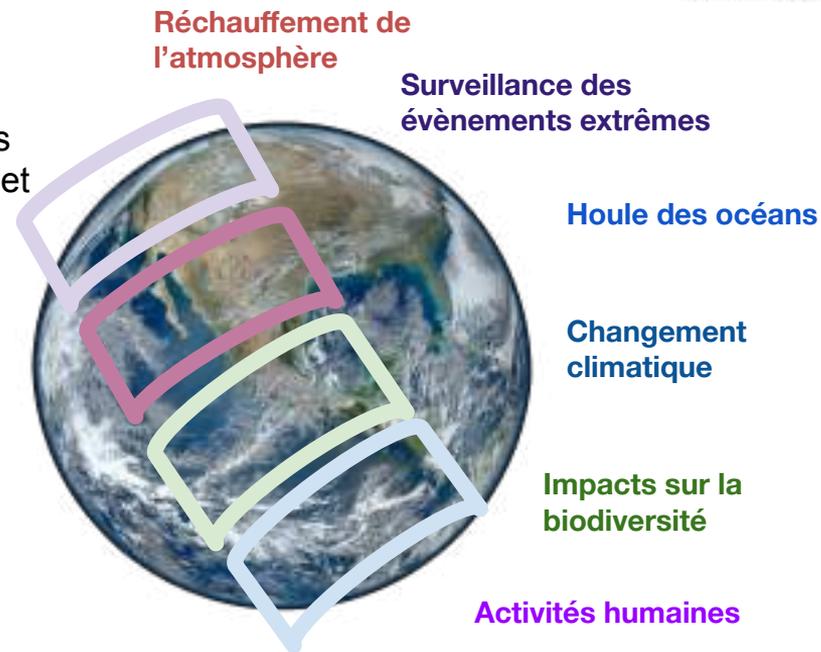


Pôle National
de Données de Biodiversité

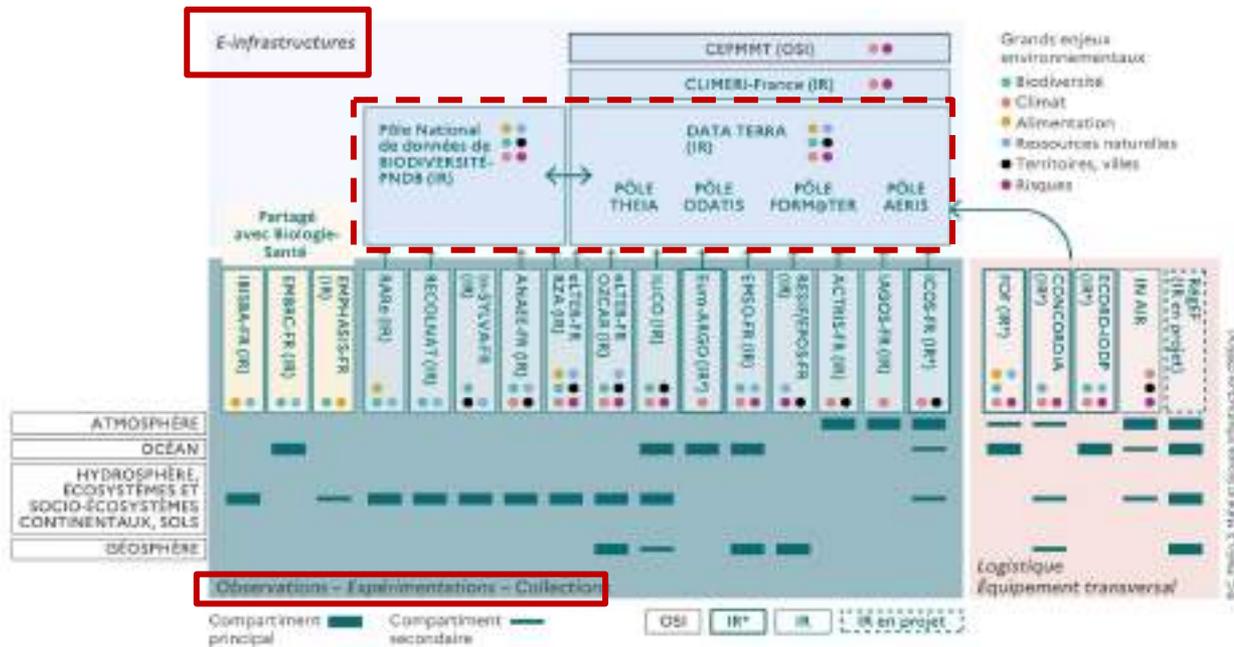
- De nombreux processus géophysiques et environnementaux,
- A différentes échelles spatiales et temporelles,
- Interactions permanentes entre la Terre solide, les surfaces continentales, l'océan, les compartiments atmosphériques et anthroposphère.

ÉVOLUTION DES BESOINS

- **Approches plus intégrées de la complexité**
- **Données multi-sources, multi-capteurs permettant une multi-échelle (in situ, spatiale, spectrale et temporelle), séries long terme.**
- **Gestion de la diversité et volume : services cloud, Big Data et IA.**
- **Interdisciplinarité et transversalité des communautés scientifiques**



PAYSAGE ACTUEL DES INFRASTRUCTURES DE RECHERCHE



PND B et DATA TERRA sont **deux E-Infrastructures de Recherche** par **Grands compartiments du système Terre** et **grands enjeux environnementaux** :

- PND B : Biodiversité
- THEIA : Surf.continentales
- ODATIS : Ocean
- FORM@TER : Terre Solide
- AERIS : Atmosphere

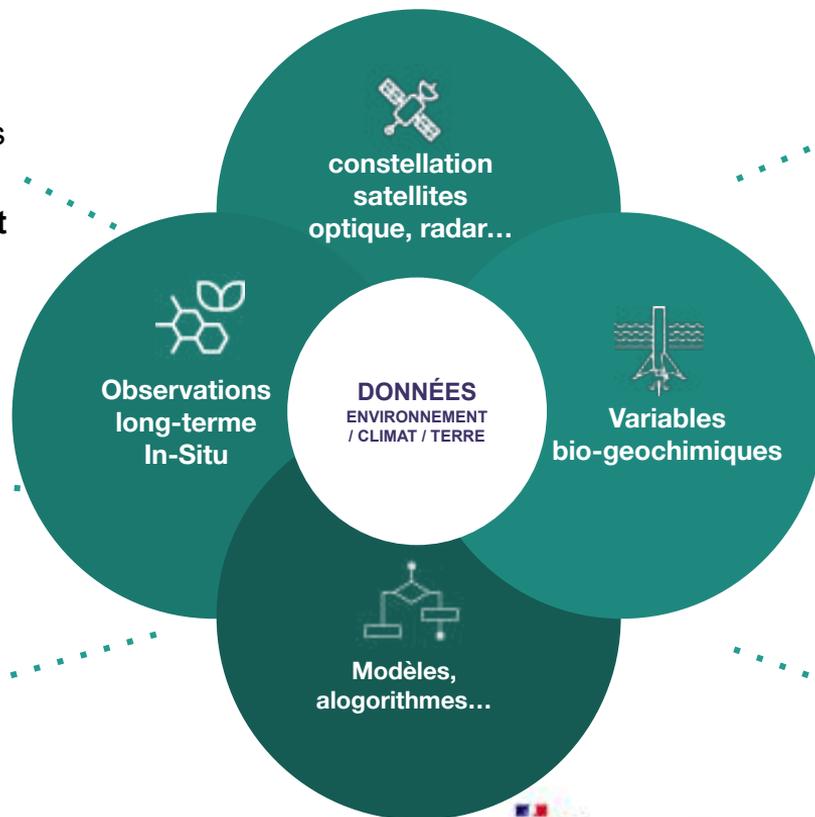
Figure 1: OSI/IR*/IR du domaine SST & ENV par grands types (observations – expérimentaux – collections, logistiques et e-infrastructures), par grands compartiments du système Terre (atmosphère, océan, hydrosphère-écosystèmes – socio-écosystèmes continentaux – sols et géosphère) et par grands enjeux environnementaux (biodiversité, climat, alimentation, ressources naturelles, territoires-villes, risques). En jaune figurent trois infrastructures partagées avec le domaine Biologie – Santé.

LA RÉPONSE DES IRs AUX DÉFIS SCIENTIFIQUES ET TECHNIQUES

Contribution aux initiatives internationales et européennes en **appui aux politiques publiques de développement durable**

Accès simplifié aux données multi-sources

Interopérabilité des services



Développement de services **sur le cycle complet de la donnée**

Se baser sur les capacités, institutions, structures et **ressources existantes**

Approches multidisciplinaires intégrées pour l'utilisation des données de recherche d'observation de la Terre

Comprendre les données par les {méta}données : définitions et contexte

"Les données de recherche sont définies comme des enregistrements factuels utilisés comme sources primaires pour la recherche scientifique, et qui sont généralement acceptés dans la communauté scientifique comme nécessaires pour valider les résultats de la recherche."

OCDE, 2007. <https://www.oecd.org/sti/inno/38500813.pdf>

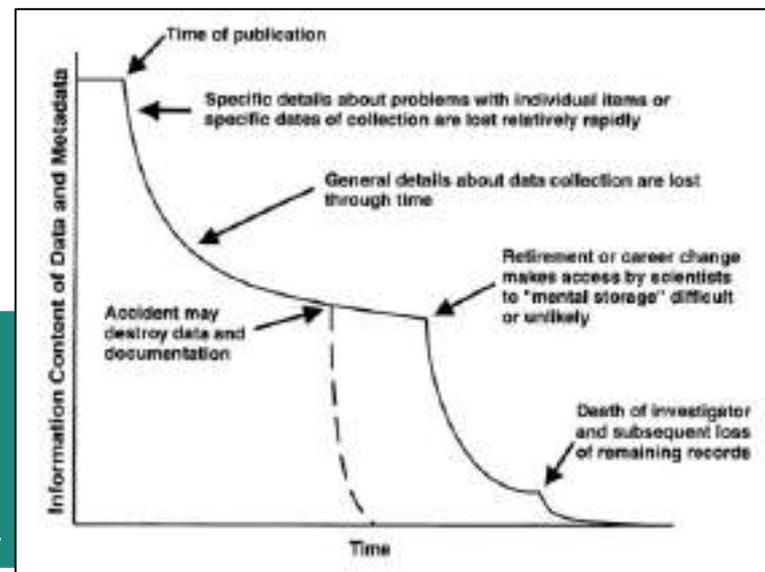
OECD Principles and Guidelines for Access to Research Data from Public Funding



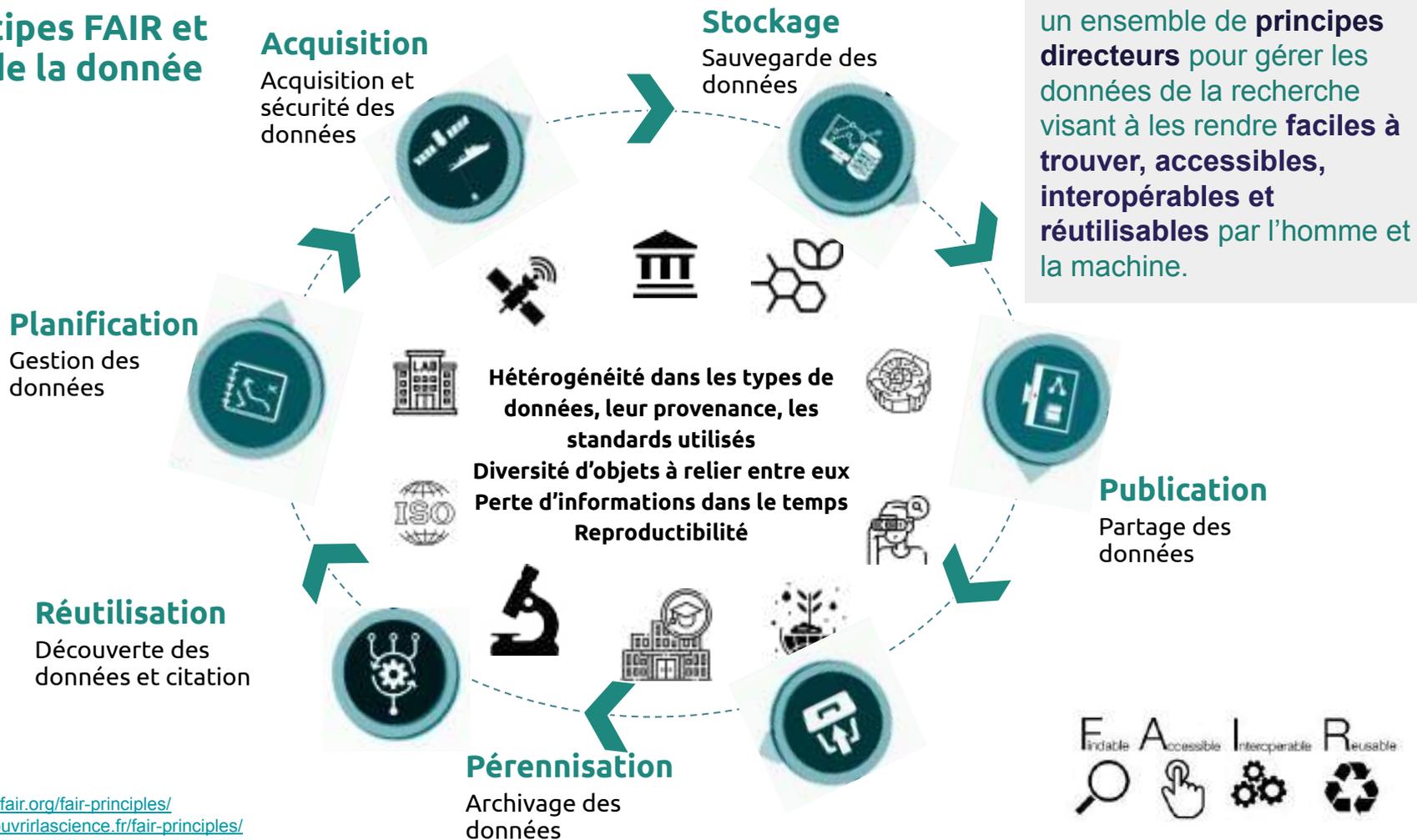
Qui ?
Quoi ?
Où
Quand ?
Comment ?

"Les métadonnées, que l'on peut définir simplement comme « des données sur les données », sont un moyen de nommer les choses et de représenter les données et leurs relations."

Christine L. Borgman, 2020. <https://books.openedition.org/oep/14692>

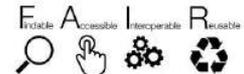


Principes FAIR et Cycle de la donnée





02 LES IRs DATA TERRA et PNDB



DATA TERRA, une e-Infrastructure de Recherche dédié au système Terre



€42m
(2020)



1000
produits &
services



15,000
usagers

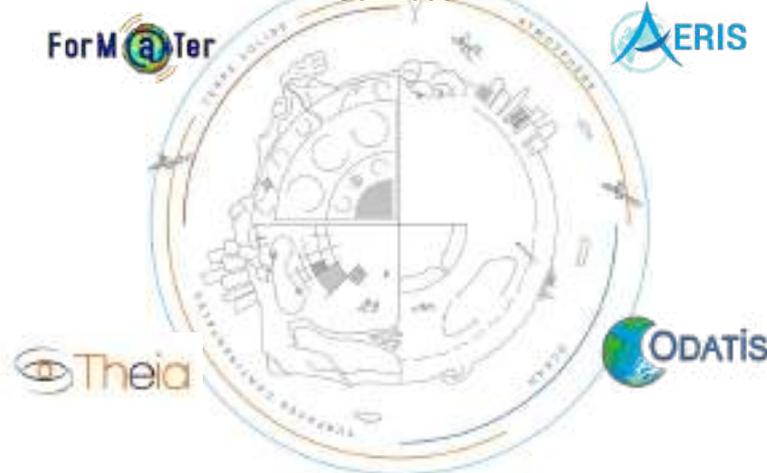


100,000 TB
(2022/2023)

La mission principale de l'IR DATA TERRA est de développer un **dispositif global d'accès et de traitement de données, produits et services** dédiés à l'**observation de la Terre**.

- 26 organismes de recherche et universités partenaires
- 30 Centre de Données et de Services
- 32 Conseils scientifiques
- 200 FTE / 450 scientifiques, ingénieurs et techniciens

4 Pôles de Données et de Services



2 Dispositifs



FAIR Data Communautés



Imagerie spatiale THRS



PNDB, le Pôle National de Données de Biodiversité



2018, Inscription sur la **feuille de route des Infrastructures** nationales de recherche du Ministère de l'Enseignement Supérieur, de la Recherche (MESR), pilotée par Allenvi.

Portage par **18 établissements partenaires**
(+ **Fondation pour la recherche sur la biodiversité, FRB**).

=> Mise en œuvre confiée au **Muséum National d'Histoire Naturelle** via PatriNat (OFB-MNHN-CNRS-IRD)
et un co-portage de l'animation avec la FRB



9 organismes et 10 universités

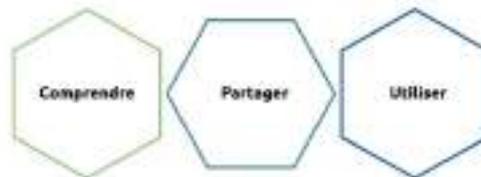


2022 : **Centre de référence thématique** "système Terre-Environnement" (MESR & Recherche Data Gov.) =>
Structuration des communautés autour des données



Horizon 2024 : 5ème pôle de l'IR Data Terra

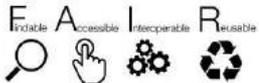
=> Une e-infrastructure ***pour et par*** les communautés de Recherche pour accompagner et faciliter la compréhension, le partage et l'utilisation des {méta}données de biodiversité



www.pndb.fr

OUTILS & SERVICES

*Hétérogénéité dans les types de données, leur provenance, les standards utilisés
Diversité d'objets à relier entre eux, Perte d'informations dans le temps, Reproductibilité*



- Partage des métadonnées
- formations
- Data paper
- Tutoriels
- DOI
- Articles & synthèses
- A.R.T.
- Support de campagne



- Catalogue de données
- Lien recherche / expertise
- Environnement virtuel de recherche
- Indicateurs
- Plan de gestion de données
- Appel à projet

Infrastructures de recherche et centres de références thématiques proposant des outils & services efficaces et complémentaires pour la **description, la mise à disposition, la validation, l'analyse, le partage et la réutilisation** des données.

ENTRÉES USAGERS (Producteurs / utilisateurs données)

- AERIS** Atmosphere
- FORM@TER** Terre solide
- ODATIS** Océans
- THEIA** Surf.continentales
- PNDB** Biodiversité
- DINAMIS** Imagerie HD

GAIA DATA, Equipex+ PIA 3 un projet structurant



DATA
TERRA



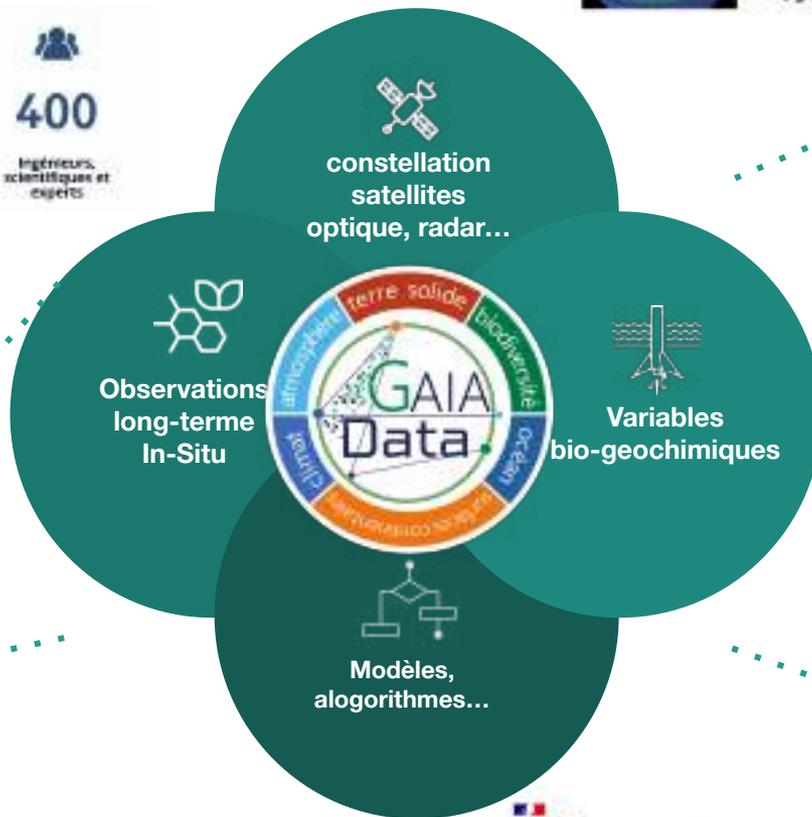
8	21	62 M €	400
Années pour la réalisation du projet	Organismes partenaires	Budget Total (dont 14,2 millions d'euros de l'ANR)	Ingénieurs, scientifiques et experts

Contribution aux initiatives internationales et européennes en **appui aux politiques publiques de développement durable**

Accès simplifié aux données multi-sources

Interopérabilité des services

www.gaia-data.org



Développement de services **sur le cycle complet de la donnée**

Se baser sur les capacités, institutions, structures et **ressources existantes**

Approches multidisciplinaires intégrées pour l'utilisation des données de recherche d'observation de la Terre

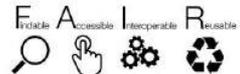


Ouvrir la science!



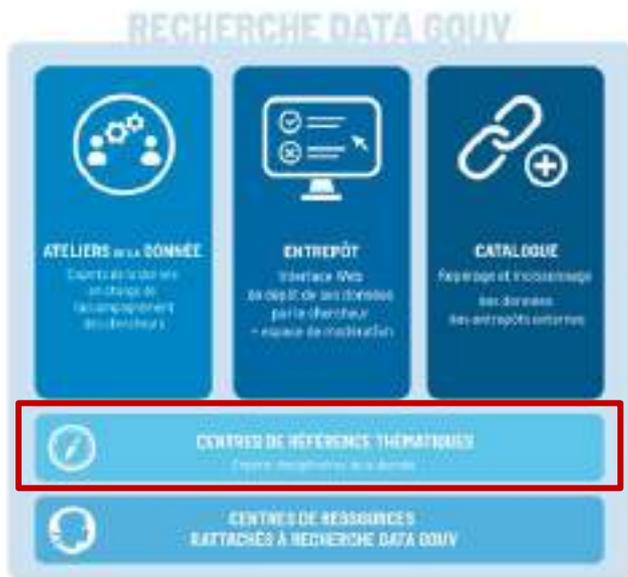


03 CENTRES DE RÉFÉRENCE THÉMATIQUES SYSTÈME TERRE ET ENVIRONNEMENT



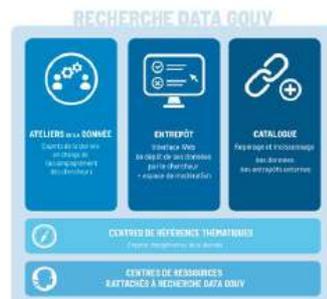
CENTRES DE RÉFÉRENCE THÉMATIQUES SYSTÈME TERRE ET ENVIRONNEMENT POUR CONTRIBUER A LA SCIENCE OUVERTE

Un écosystème au service du partage et de l'ouverture des données de recherche
FÉDÉRER, ACCOMPAGNER, PARTAGER, OUVRIR, RÉUTILISER



Structuration progressive des centres de références thématiques avec les organismes de Recherche : **DATA TERRA** et **PNDB** dans le domaine **Système Terre en Environnement**

CENTRES DE RÉFÉRENCE THÉMATIQUES SYSTÈME TERRE ET ENVIRONNEMENT POUR CONTRIBUER A LA SCIENCE OUVERTE



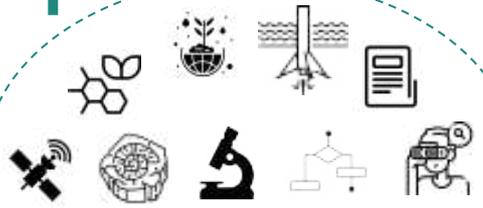
Définir

- les **normes internationales** de description et de diffusion des données (ouverture, période d'embargo, accès restreint).
- et **diffuser les bonnes pratiques** de collecte, documentation, traitement, et diffusion des données.
- la **liste des entrepôts de données de référence** (nationaux et/ou internationaux) vers lesquels orienter les chercheurs pour le dépôt des données et que *Recherche Data Gov* moissonnera

Contribuer à la définition de l'**arborescence** des données de l'entrepôt *Recherche Data Gov* **thématiques Système Terre et Environnement**

Soutenir l'**articulation** entre les **dispositifs thématiques** spécialisés et *Recherche Data Gov* et l'**inscription** dans l'**écosystème international**

1



données Terre-Environnement :
in situ, données satellitaires,
modèles algorithmiques...

Caractérisation de mon jeu de donnée (=métadonnées) via un accompagnement

- Qui est le "data owner" ?
- De quoi parle mon jeu de données ?
- Quel est le cadre d'acquisition (recherche, politiques Publiques, ...) ?
- Quel est le format et standard utilisé ?
- Publication d'un data paper en parallèle ?
- création d'un Plan de gestion de données

2



exemples de Systèmes d'Informations
dont entrepôts institutionnels et
catalogues de données des
infrastructures

Données ouvertes : moissonnage, standardisation
des {méta}données et enrichissement (description
attributs et termes issus de ressources
terminologiques) **des métadonnées** + liens directs
de téléchargement

Centres de références
Thématiques

Carte des
ateliers de
la donnée

Centres de ressources

PNDDB
Pôle National
de Données de Biodiversité

DATA
TERRA

outils et services pour la
description, l'accès, la validation,
l'analyse et la réutilisation des
données Terre-Environnement

4



Systèmes d'informations et
catalogues de données
internationaux

Métadonnées et workflows analytiques visibles sur
catalogues/instances internationales

3

Infrastructures de recherche

DATA
TERRA

PNDDB

PARTENAIRES DATA TERRA



PARTENAIRES PNDB





DATA
TERRA



PNDB

CENTRES DE RÉFÉRENCES THÉMATIQUES
Système Terre-Environnement

Merci ! Questions ?

Love Data Week
Nancy, Université de Lorraine, 16 mars 2023

Par
Ghislaine Abbassi, responsable Communication Data Terra
et
Olivier Norvez, responsable Animation du PNDB



Idées de réflexions / échanges ?!

Où mettre / partager vos données ?

Quelles sont les licences d'utilisation ?

*Quelles sont les différences et complémentarité entre centre de ref thématique /
centre de ressources / ateliers de la donnée*

Que comprend le système Terre-Environnement ?

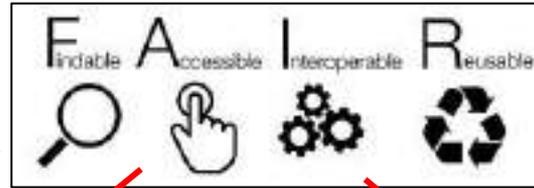
Quelles sont vos attentes (types d'actions, documentations, guides, etc.) ?

Prochain RDV Data Terra / PNDB avec les ateliers de la donnée 27 MARS

? Idées

- Où mettre / partager vos données ?
 - *entrepôts institutionnels, entrepôts thématiques*
- connaissances sur les licences d'utilisation ?
 - *CC-BY 2.0*
- différences et complémentarité entre centre de ref thématique / centre de ressources / ateliers de la donnée
 - *???*
- Fonctionnement de Recherche data gouv
 - *bottom up / co-construction / se base sur de l'existant (outils, RH) pour mieux orienter/informer/former/coordonner les initiatives locales et globales*
- Que comprend le système terre-environnement ?
 - *géologie, océanographie, biodiversité, atmosphère, surfaces continentales, etc.*
- types d'actions demandées / attendues par les communautés
 - *session data paper, atelier FAIR Implementation Profile,*

Des outils et services adaptés : *l'approche FAIR*



Facile à trouver

Métadonnées détaillées dont renseignements des attributs

Accessible

Licence ouverte, téléchargement direct des données

Interopérable

Standards internationaux, ontologies, thésaurus, outils open source

Réutilisable

Réutilisation par mon "futur moi" et par les autres

ATMOSPHERE

Physique et chimie de l'atmosphère, changement climatique, qualité de l'air, etc

Le pôle de données et services AERIS regroupe les activités de **gestion des données** et **l'expertise scientifique de l'atmosphère**.

Les données proviennent de plateformes, de réseaux d'observation, d'instruments à bord de satellites, de ballons, d'avions, de mesures en laboratoire, d'inventaires, de modèles, de données de campagnes scientifiques.

ETUDE DE CAS : QUALITÉ DE L'AIR

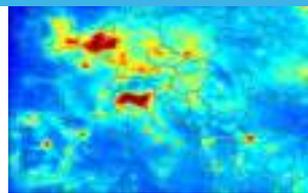
SURVEILLANCE QUALITÉ DE L'AIR



ESTIMATIONS DES ÉMISSIONS DE CARBONE



RÉSEAU EUROPÉEN DE LA QUALITÉ DE L'AIR



EFFETS DES FEUX DE FORÊTS SUR LA QUALITÉ DE L'AIR



www.aeris-data.fr/



Connaissance de la structure et de la composition de la Terre, risques et ressources telluriques

Le pôle de données et services FORMATER regroupe les activités de **gestion de données** et **l'expertise scientifique de la terre solide**.

Les données proviennent de **missions satellitaires**, **d'instruments in situ** et **des résultats d'expériences**, **de modélisations** et **de simulations**. Des services de calcul sont également disponibles.

ETUDE DE CAS

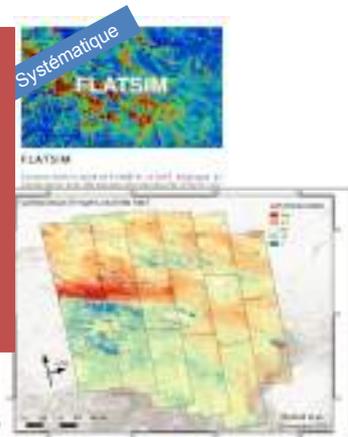
CALCUL DE LA DEFORMATION DU SOL PAR IMAGERIE SPATIALE

SURVEILLANCE SPATIALE TEMPORELLE DES RÉGIONS CRITIQUES

APPLICATIONS

Zones de failles actives et systèmes magmatiques, glissements de terrain et déformations lithosphériques à l'échelle des grands massifs.

A la demande



Observations et compréhension des océans au service du développement durable

Le pôle de données et services OCEAN regroupe les activités de **gestion de données** et de **produits pour le bénéfice des sciences océaniques**

Les données proviennent de **missions satellitaires**, **d'instruments in situ** (flotteurs, planeurs, radars, campagnes en mer), **d'expérimentations** et de **mesures en laboratoire**.



DIVERSITE DES DONNEES : de la physique à la biologie

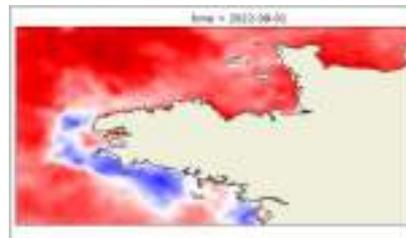
SURVEILLANCE DU
LITTORAL



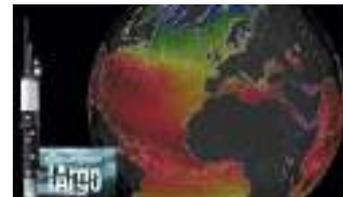
QUALITE DES EAUX /
HYPOXIE



TEMPÉRATURE, VAGUE
DE CHALEUR MARINE



BIOGÉOCHIMIE



www.odatis-ocean.fr

Observation et compréhension des dynamiques des surfaces terrestres et des ressources naturelles

Le pôle de données et services THEIA regroupe les activités de **gestion de données** et **l'expertise scientifique des surfaces continentales**.

Les données proviennent de **missions satellitaires, de réseaux d'observations et d'instruments in situ**. Il s'agit d'accompagner le développement de produits à valeur ajoutée et de services de traitement sur différentes thématiques (eau, forêt, neige & glace, risques naturels, santé, occupation des sols, urbain, agriculture, biodiversité, littoral, etc.).

ÉTUDE DE CAS

OCCUPATION DES SOLS



TACHE ARTIFICIALISEE



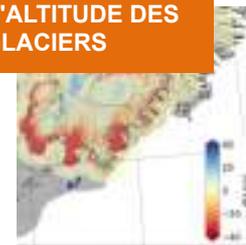
PRODUCTION DE LA BIOMASSE



HUMIDITÉ DES SOLS



CHANGEMENT D'ALTITUDE DES GLACIERS



COUVERTURE DE NEIGE



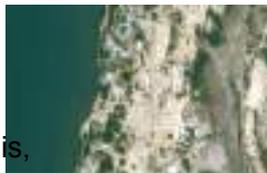
DISPOSITIF DINAMIS

Accès et utilisation d'images satellitaires à très haute résolution spatiale

UNE OFFRE INSTITUTIONNELLE

Couvertures Pléiades

les littoraux sablonneux métropolitains et guyanais, DOM-TOM...



Littoral Sénégal. Pléiades
Copyright CNES, ADS



Montpellier. Pléiades © CNES 2020, distribution Airbus DS

Couvertures SPOT 6-7 France métropolitaine

Acquisitions ponctuelles Pléiades et SPOT 6-7
Monde entier - Exprimé par les AIU, quel que soit leur lieu d'affectation.



Couverture métropolitaine 2020. Copyright ADS - IRD, INRAE, IGN

Images complémentaires haute résolution

Relais programme CNES Spot patrimoine mondial (Spot 1-5), plateforme CNES PEPS (Sentinelle 2), Geosud, programme CNES Kalideos : RapidEye, CosmoSkyMed, TerraSar, -X, Aster...

3 TYPES D'USAGERS sous condition

Utilisateurs institutionnels français

Pour la recherche scientifique, l'enseignement supérieur et les acteurs territoriaux

Utilisateurs institutionnels et scientifiques étrangers

Dans le cadre de coopération au niveau européen et international

Utilisateurs privés (France)

Dans le cadre de projets R&D et de prestations de la commande publique

dinamis.data-terra.org



Le Pôle National de Données de Biodiversité : *préambule*

Le PNDB, pour quoi faire ?

Données de recherche

- Observation et suivis au sol (y compris sc. participatives) ou satellitaires,
- Collections naturalistes
- Séquençage, expérimentation, imagerie...
- Données secondaires issues de la modélisation



Entrepôts de données (et SI)

- Organismes, universités.

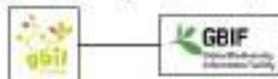


- Infrastructures du système "Terre-Environnement" et leurs miroirs européens



Etc.

- Autres producteurs (ex. microorganismes pathogènes; data.gouv, GBIF...)



Infrastructures numériques

Pôles de données



FAIRisation
Structuration
des communautés
par les données



Utilisateurs / Réutilisateurs

- Existent ?
- Localisation ?
- Accès ?
- Description ?
- Interopérabilité ?
- Croisement de jeux de données



Une infrastructure pour et par la Recherche : Gouvernance et interactions

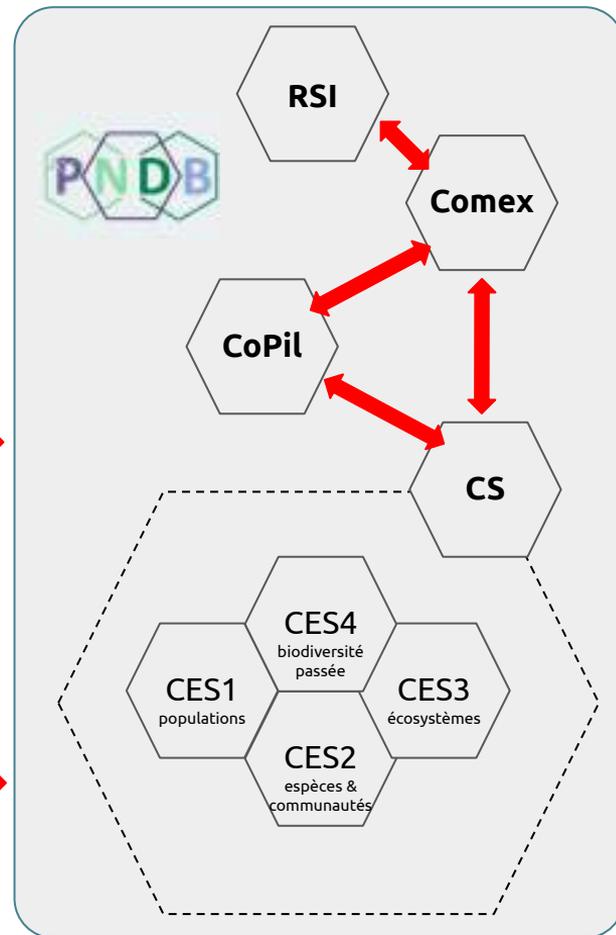
La gouvernance du PNDB est composée de 4 instances : Le **Comité de pilotage** (CoPil), le **Conseil Scientifique** (CS) et ses Comités d'Experts Scientifiques (CES), le **Comité Exécutif** (Comex) et le **Réseau des responsables des systèmes d'information** (RSI).



National



International



Une infrastructure pour et par la Recherche : Le Conseil scientifique

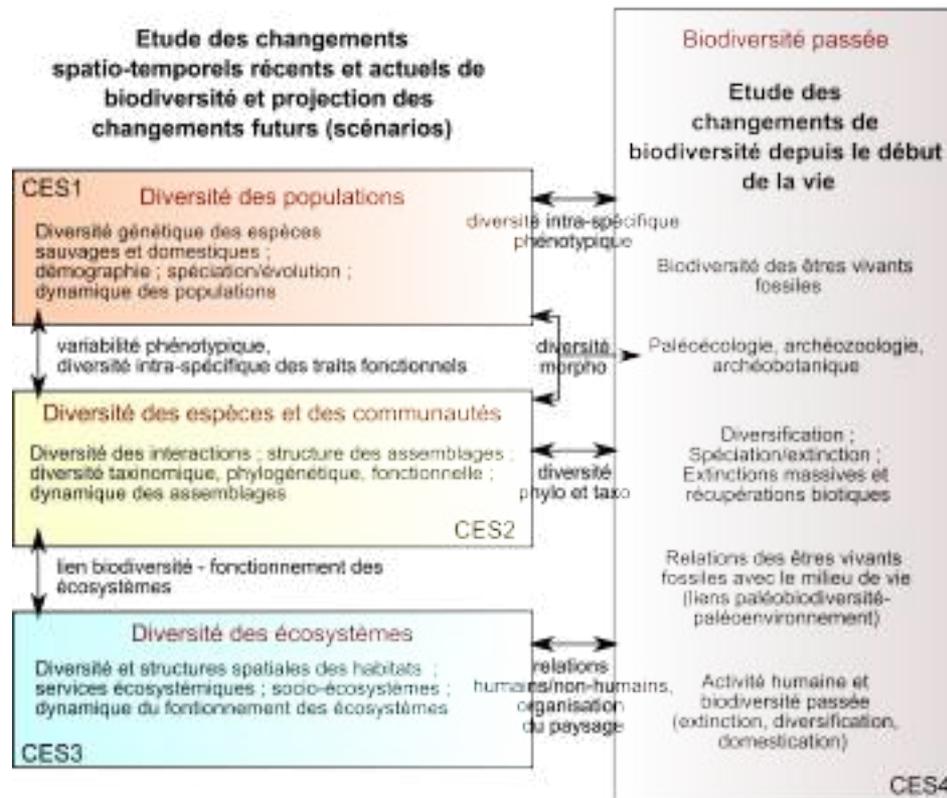
Compréhension des dynamiques passées, récentes et actuelles de biodiversité et projection des dynamiques futures à toutes les échelles spatiales

Le CS émet des avis et des recommandations concernant :

- les orientations,
- les stratégies
- et les livrables scientifiques du PNDB, y compris en ce qui concerne les projets soumis à appels à projets financés .

Ses avis sont communiqués au Copil et au Comex.

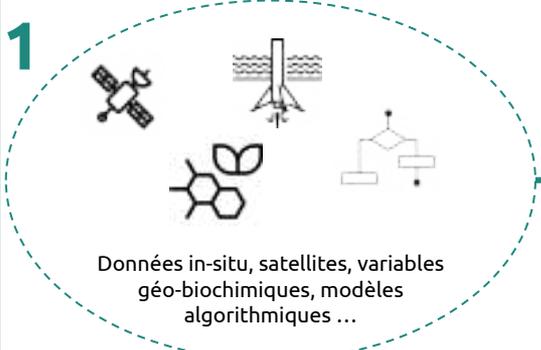
Pour ce faire, Le PNDB compte quatre **Comités d'experts scientifiques (CES)** dont les thèmes et responsables sont les suivants :



Cycle de la donnée biodiversité



1



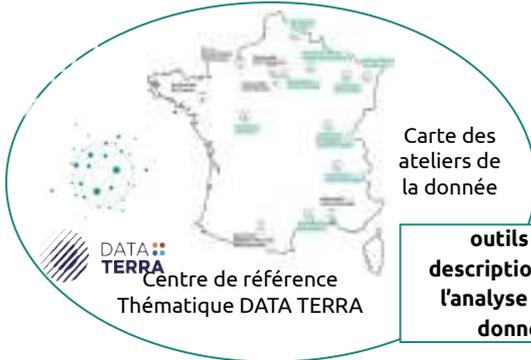
Caractérisation de mon jeu de donnée (=métadonnées) via un accompagnement

- Qui est le "data owner" ?
- De quoi parle mon jeu de données ?
- Quel est le cadre d'acquisition (recherche, politiques Publiques, ...) ?
- Quel est le format et standard utilisé ?
- Publication d'un data paper en parallèle ?
- création d'un Plan de gestion de données

2

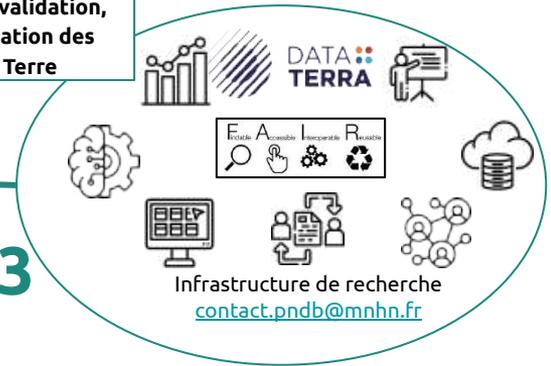


Données ouvertes : moissonnage, conversion en EML et enrichissement (description attributs et termes issus de ressources terminologiques) des métadonnées + liens directs de téléchargement



outils et services pour la description, l'accès, la validation, l'analyse et la réutilisation des données système Terre

3



Métadonnées et workflows analytiques visibles sur catalogues/instances internationales

4



1

Séquences génétiques, occurrences d'espèces, liste sp, fossiles, données satellites, flux de carbone...

Caractérisation de mon jeu de donnée (=métadonnées) via un accompagnement

- Qui est le "data owner" ?
- De quoi parle mon jeu de données ?
- Quel est le cadre d'acquisition (recherche, politiques Publiques, ...) ?
- Quel est le format et standard utilisé ?
- Publication d'un data paper en parallèle ?
- création d'un Plan de gestion de données

2

exemples de Systèmes d'Informations de biodiversité dont entrepôts institutionnels et catalogues de données des infrastructures (RSI)

Données ouvertes : moissonnage, conversion en EML et enrichissement (description attributs et termes issus de ressources terminologiques) des métadonnées + liens directs de téléchargement

Carte des ateliers de la donnée

PNDB Centre de référence Thématique (PNDB)
contact.pndb@mnhn.fr

outils et services pour la description, l'accès, la validation, l'analyse et la réutilisation des données de biodiversité

Métadonnées et workflows analytiques visibles sur catalogues/instances internationales

4

Systèmes d'informations et catalogues de données internationaux

3

Infrastructure de recherche
contact.pndb@mnhn.fr